

## PUTTING BEHAVIORAL ECONOMICS TO WORK: TESTING FOR GIFT EXCHANGE IN LABOR MARKETS USING FIELD EXPERIMENTS

BY URI GNEEZY AND JOHN A. LIST<sup>1</sup>

Recent discoveries in behavioral economics have led scholars to question the underpinnings of neoclassical economics. We use insights gained from one of the most influential lines of behavioral research—gift exchange—in an attempt to maximize worker effort in two quite distinct tasks: data entry for a university library and door-to-door fundraising for a research center. In support of the received literature, our field evidence suggests that worker effort in the first few hours on the job is considerably higher in the “gift” treatment than in the “nongift” treatment. After the initial few hours, however, no difference in outcomes is observed, and overall the gift treatment yielded inferior aggregate outcomes for the employer: with the same budget we would have logged more data for our library and raised more money for our research center by using the market-clearing wage rather than by trying to induce greater effort with a gift of higher wages.

KEYWORDS: Gift exchange, field experiment.

### 1. INTRODUCTION

NEOCLASSICAL ECONOMICS treats labor as a hired input in much the same manner as capital. Accordingly, in equilibrium, the firm pays market-clearing wages and workers provide minimum effort. The validity of this assumption is not always supported by real-life observations: some employers pay more than the market-clearing wage and workers seemingly invest more effort than necessary (Akerlof (1982)). The “fair wage–effort” hypothesis of Akerlof (1982) and Akerlof and Yellen (1988, 1990) extends the neoclassical model to explain higher than market-clearing wages by using a gift exchange model, where “On the worker’s side, the ‘gift’ given is work in excess of the minimum work standard; and on the firm’s side the ‘gift’ given is wages in excess of what these women could receive if they left their current jobs” (Akerlof (1982, p. 544)).<sup>2</sup>

The gift exchange model is based on the critical assumption of a positive relationship between wages and worker effort levels. Workers are assumed to respond to high wage levels by increasing their effort (positive reciprocity) and to low wage levels by decreasing their effort (negative reciprocity) to the minimum required, in retaliation for the low wage. A large and influential body

<sup>1</sup>A co-editor and four anonymous reviewers provided comments that significantly improved the study. Ernst Fehr and Glenn Harrison also provided remarks that improved the manuscript. Stefan Andersen provided fine research assistance. Craig Landry, Andreas Lange, Michael Price and Nicholas Rupp also helped with gathering the data. Jamie Brown-Kruse worked with us in her capacity as the Director of the Hazards Center.

<sup>2</sup>The notion of gift exchange was apparently first proposed by Adams (1963), who posited that in social exchange between two agents the ratio of the perceived value of the inputs (e.g., wage) to the perceived value of outputs (e.g., resulting from the employee’s effort) would be equivalent.

of experimental evidence in support of reciprocity has been reported in the past two decades.<sup>3</sup> One of the first experiments to test this assumption is Fehr, Kirchsteiger, and Riedl (1993), who constructed a market with excess supply of labor, ensuring a low equilibrium wage. Under their setup, employees had no pecuniary incentive to raise the quality of their work above the exogenously given minimum. If an employer expects employees to invest only the minimum effort required, then she has no compelling reason to pay wages above the market-clearing level. Contrary to this prediction, however, most employers attempted to induce employees to invest greater effort by offering them higher (at times by more than 100%) than market-clearing wages. On average, this high wage was reciprocated by greater employee effort, making it profitable for employers to offer high wage contracts. Subsequent laboratory exercises have largely led to similar conclusions.

Whereas the literature has taken the experimental results as providing key support for the received labor market predictions of Akerlof (1982) (see, e.g., Fehr, Kirchsteiger, and Riedl (1993, p. 437), who note that their results “provide... experimental support for the fair wage–effort theory of involuntary unemployment”), it remains largely unknown whether such inference is appropriate from observed laboratory behavior. In this case, for example, is the behavior of laboratory subjects, who are asked to choose an effort or wage level (by circling or jotting down a number) in response to pecuniary incentive structures, a good indicator of actual behavior in labor markets?

We tackle this issue directly by focusing on real effort in labor markets using field experiments. In this regard, one key missing link in the literature between the laboratory and labor markets is the *duration* of the task. Whereas interaction in the lab is typically abbreviated and usually takes no longer than one or two hours, interaction in labor markets typically lasts weeks, months, or years. One lesson learned from the psychology literature is that there are important behavioral differences between psychological processes in the short run and in the long run: for example, *hot* versus *cold* decision making. In many cases, the immediate reaction to an event is dictated by what is called hot decision making (Loewenstein (2005)). People’s decision making behavior in this hot phase is different than in the “cold” phase (the folk wisdom of counting to 10 before one reacts is based on such a difference). Additionally, *adaptation* has been found to be important. Much like the human eye adapts to changes

<sup>3</sup>See, for example, Kahneman, Knetsch, and Thaler (1986), Fehr, Kirchsteiger, and Riedl (1993), Fehr, Gächter, and Kirchsteiger (1997), Berg, Dickhaut, and McCabe (1995), Guth, Schmittberger, and Schwarze (1982), Guth (1995), Roth (1995), Charness (2004), Falk and Gächter (2002), and Fehr and Falk (1999). For a survey and discussion of positive and negative reciprocity, see Fehr and Gächter (2000). For a survey of the theoretical models of reciprocity, see Sobel (2005). This picture is confirmed by scholarly work in many other fields. It is accounted for in interview studies that economists have conducted with business leaders (Agell and Lundborg (1995)) and it is consistent with discussions in organization theory (Steers and Porter (1991)) and psychology (Argyle (1981)).

in light, decision makers tend to adapt to new situations over time (Gilbert, Pinel, Wilson, Blumberg, and Wheatley (1998)).

To provide a first test of the gift exchange hypothesis in an actual labor market, we invited people to take part in an effort to computerize the holdings of a small library at a large university in the Midwest. Recruitment was done via posters that promised participants *one-time* work that would last six hours and that would pay \$12 per hour, or \$72. Participants were not informed that they were taking part in an experiment.

The first treatment paid a flat wage of \$12 per hour, as promised. In the second treatment, once the task was explained to the participants, they were told that they would be paid \$20 per hour, not the \$12 that had been promised. We found that, in line with the gift exchange hypothesis, participants in the \$20 treatment provided significantly higher effort in the first 90 minutes than participants in the \$12 treatment. After 90 minutes on the job, however, effort levels were indistinguishable across the two treatments.

Our second field experiment invited students to take part in a door-to-door fundraising drive to support the Natural Hazards Mitigation Research Center at a large university in the Southeast. Similar to the library task, participants were told that this was *one-time* work for which they would be paid \$10 per hour. An important difference in this case is that workers have a better idea about the surplus and how much the employer valued the task. This difference is important because if the surplus is known, the share of the surplus that the workers receive will determine whether they perceive their wage as fair.<sup>4</sup>

The first treatment was a flat wage of \$10 per hour, as promised. In the second treatment, once the solicitors were trained, they were told that they would be paid \$20 per hour, not the \$10 that had been advertised. Empirical results mirror those in the library task: solicitors in the \$20 treatment raised significantly more money in the first few hours of the task than solicitors in the \$10 treatment, but after a few hours the observed outcomes were indistinguishable.

The remainder of our study proceeds as follows. Section 2 presents the experimental design. Section 3 discusses the experimental results. Section 4 summarizes how our work relates to the literature. Section 5 concludes.

## 2. EXPERIMENTAL DESIGNS

### *A. Library Task*

Our first field experiment was conducted (by Gneezy) at a large university, using undergraduate student participants who were invited to take part in an effort to computerize the holdings of a small library at the university. Recruitment posters informed potential participants that this was *one-time* work that

<sup>4</sup>Alternatively, if workers know only the promised wage and not the surplus, as in our library task, only the promised wage can serve as a reference point. We thank Ernst Fehr for pointing us in this direction.

would last exactly six hours, for which they would be paid \$12 per hour. Those interested in participating were instructed to call a phone number during specified times. The number connected callers to a research assistant who explained particulars about the task and payment. For those interested, the research assistant then scheduled a time in which they could arrive and perform the task. The same research assistant was used for all callers and participants. The research assistant knew none of the participants personally.

Treatment *noGift* offered laborers a flat wage of \$12 per hour, as promised. In the second treatment, denoted treatment *Gift*, once the task was explained to the participants they were told that they would be paid \$20 per hour rather than the \$12 rate advertised.

### *The task*

Participants were seated in front of a computer terminal next to boxes filled with books and were asked to enter data regarding the books into a data base on the computer. The data included title, author, publisher, ISBN number, and year of publication. Each participant performed the task alone, without viewing the other participants. Participants could take a break from their work whenever necessary. The experimental monitor recorded the number of books they entered every 90 minutes.

Before proceeding to a discussion of the second field experiment, we should briefly mention a few design issues. First, students were not informed that they were taking part in an experiment. This is important given that we wished to observe subjects in a natural work environment. Second, we were careful to remind them that this was a one-time work opportunity. This is important in light of an alternative theory to the fair wage–effort hypothesis—efficiency wage theory—which surmises that employers pay above market-clearing wages to motivate workers to increase their effort level so as to avoid being fired, which economizes on firm-level monitoring (see, e.g., Katz (1986)). Third, we calculated that we would need roughly 100–120 hours of total worker time to complete the task. Thus, to test our duration hypothesis we hired 19 workers for six hours each, splitting the sample into  $n = 10$  and  $n = 9$  for the *noGift* and *Gift* treatments, respectively. Finally, we distributed books across individuals randomly to ensure orthogonality of book type and treatment.

### *B. Fundraising Task*

Our second field experiment was part of a door-to-door fundraising drive to support the Natural Hazards Mitigation Research Center (henceforth the Hazards Center) at a large university.<sup>5</sup> Door-to-door fundraising is widely used

<sup>5</sup>The Natural Hazard Mitigation Research Center was authorized to begin operations in the fall of 2004 by the North Carolina state government. The Hazard Center was founded in response to

by a diverse range of organizations. In both treatments, which follow the two library task treatments, households in predetermined neighborhood blocks in Pitt County, North Carolina, were approached by a paid solicitor and asked if they would like to make a contribution to support the Hazards Center. Households where someone answered the door were provided an informational brochure about the Hazards Center and were read a fixed script that outlined the reason for the solicitor's visit. The script included a brief introduction that informed residents who the solicitors were, the purpose of their visit, and a one- to two- sentence summary of the nonprofit organization. Copies of the script are provided in the [Appendix](#).

As summarized in the [Appendix](#), potential donors were informed that all proceeds raised in the fundraising campaign would be used to fund the Hazards Center. Households were also informed that each dollar contributed to the Hazards Center would provide them with one ticket for a \$1,000 lottery, where the chances of winning the prize were based on the total number of tickets allocated.<sup>6</sup>

The experimental treatments were conducted on two different weekends in November 2004. In total, we employed 23 solicitors—10 in the *noGift* treatment and 13 in the *Gift* treatment. Potential subjects were recruited via flyers posted around campus, announcements on a university electronic bulletin board, advertisements in the local campus newspaper, and direct appeal to students during undergraduate economics courses. All solicitors were told they would be paid \$10 per hour during training and employment.

Each solicitor's experience typically followed four steps: (1) consideration of an invitation to work as a paid volunteer for the research center, (2) an in-person interview, (3) a training session, and (4) participation as a solicitor in the door-to-door campaign. Upon being hired (all applicants were hired), all solicitors attended a one-hour training session on Friday afternoon of the weekend they were scheduled to work.<sup>7</sup> Upon arriving on Saturday morning,

---

the widespread devastation in eastern North Carolina caused by hurricanes Dennis and Floyd, and was designed to provide support and coordination for research on natural hazard risks. For more information on the Hazard Mitigation Research Center, see [www.artsci.ecu.edu/cas/auxiliary/hazardcenter/home.htm](http://www.artsci.ecu.edu/cas/auxiliary/hazardcenter/home.htm). The design discussion follows Landry, Lange, List, Price, and Rupp (2006), who explore various mechanisms (voluntary contributions mechanisms with and without seed money, and two types of lotteries) for inducing charitable contributions.

<sup>6</sup>We also randomly placed solicitors in a treatment with a multiple-prize lottery. The data across these two lottery types are not significantly different, so we pool them and suppress further discussion. See Landry, Lange, List, Price, and Rupp (2006) for a further discussion of the broader lottery results.

<sup>7</sup>Each training session was conducted by the same researcher and covered a single treatment. The training sessions provided the solicitor with background/historical information about the Hazards Center and reviewed the organization's mission statement and purpose. Solicitors were provided a copy of the brochure and the press release announcing the formation of the Hazards Center. Once solicitors were familiarized with the Hazards Center, the trainer reviewed the data collection procedures.

subjects were split randomly into two groups, after which one group (treatment *Gift*) was informed separately that its members would receive \$20 per hour, rather than the \$10 hourly rate the other group (treatment *noGift*) received. All solicitors participated during a single weekend and were not informed that they were participants in an experimental study; neither were they informed that different solicitors were being paid different amounts. Further care was taken to keep solicitors in different experimental treatments isolated from one another to prevent cross-contamination and information exchange across treatments.

A few important design issues should be discussed before proceeding to the results summary. First, solicitors were provided with a record sheet that included columns to record the race, gender, and approximate age of potential donors, along with their contribution level. The trainer stressed the importance of recording the contribution (or noncontribution) data immediately following the solicitation “sales pitch.” This permits us to examine the temporal nature of effort and contributions secured. Second, solicitors were instructed to distribute an information brochure after introducing themselves to potential donors. This provided legitimacy to the fundraising drive, because brochures are a common tool in the industry. Finally, solicitors were instructed to wear khaki pants (or shorts) and were provided with a polo shirt that displayed the name of the Hazard Center.

### 3. EMPIRICAL RESULTS

A quick summary of the empirical results is that, consistent with the bulk of past experimental evidence from the lab, there are signs of significant gift exchange in the data: in the early hours of the task, effort (money raised) in the *Gift* treatments is markedly higher than in the *noGift* treatments. Across both tasks, however, this increased effort (money raised) wanes quickly: after the first few hours, effort levels (money raised) across the *Gift* and *noGift* treatments are statistically indistinguishable. This drop in effort (money raised) causes us to conclude that for the wage levels considered in our treatments, our resources would have been better spent hiring agents at market-clearing wages. Evidence for these empirical findings are described more fully below.

#### *A. Library Task*

Table I provides a raw data summary and Figure 1 summarizes the temporal work effort in the library task split into 90-minute intervals. In the first 90 minutes, the average number of books logged into the computer is quite different across the two treatments: whereas workers logged, on average, 51.7 books in the *Gift* treatment, an average of only 40.7 books were logged in the *noGift* treatment. This nearly 25 percent difference is statistically significant at the  $p < 0.05$  level using a one-tailed Wilcoxon (Mann–Whitney)

TABLE I  
SUMMARY DATA—BOOKS LOGGED

	Participant Number	90 Minutes	180 Minutes	270 Minutes	360 Minutes
<i>noGift</i>	1	56	61	58	63
	2	52	52	51	45
	3	46	44	52	42
	4	45	41	43	38
	5	41	29	33	25
	6	38	42	44	46
	7	37	39	38	38
	8	34	35	32	37
	9	32	32	28	27
	10	26	30	33	35
	Average	40.7	40.5	41.2	39.6
<i>Gift</i>	11	75	71	60	58
	12	64	65	63	61
	13	63	65	59	63
	14	58	40	35	31
	15	54	42	33	34
	16	47	35	28	25
	17	42	37	47	39
	18	37	29	30	30
	19	25	20	20	22
	Average	51.7	44.9	41.7	40.3

nonparametric test, which has an alternative hypothesis of the *Gift* treatment having more books logged. To construct the test statistic, we first calculated the individual mean books logged in the 90-minute period and then ranked

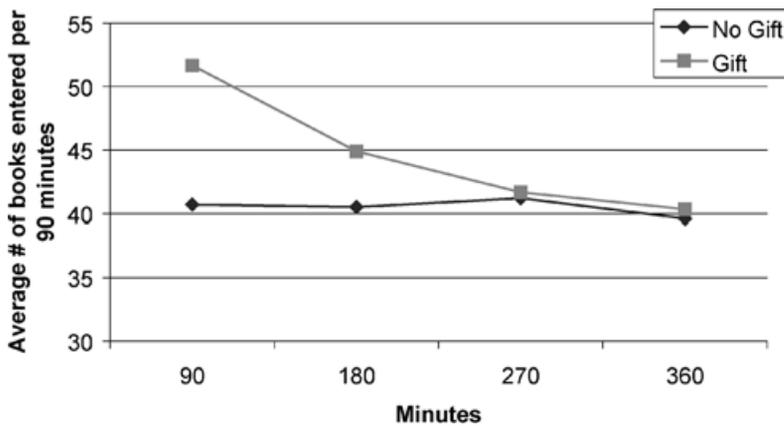


FIGURE 1.—Average books logged per time period.

subjects via these means. The test statistic is normally distributed and takes on a value of  $z = 1.76$ . A  $t$ -test assuming unequal variances also yields statistical significance at the  $p < 0.05$  level using a one-sided alternative:  $t = 1.85$ .

Although this 25 percent difference in effort is indeed noteworthy, inspection of the remainder of the temporal effort profile does not provide compelling evidence in favor of the gift exchange hypothesis. None of the remaining effort levels is significantly different at conventional levels using Wilcoxon nonparametric tests ( $z = 0.37$ ;  $z = 0.04$ ;  $z = -0.41$ ), and the data become quite similar in the final three hours;  $t$ -tests assuming unequal variances yield similar insights.

Data in Table I makes it clear that this behavioral phenomenon is not restricted to a few select workers. For example, in the *Gift* treatment every worker but one decreased their effort from the first to the last time block (the worker who did not, participant #13, had the identical number of books logged). Alternatively, in the *noGift* treatment, half of the workers increased their effort whereas half decreased their effort. Using both parametric and nonparametric matched pair's tests indicates that workers in the *Gift* treatment significantly reduced their number of books logged whereas there is no significant change in the number of books logged by workers in the *noGift* treatment. To complement these insights we estimate a panel data regression model in which we regress the individual number of books logged on a dummy variable for treatment, dummy variables for time indicators, and their interaction. Because the treatment dummy variable is static, we report panel data estimates from a random effects regression model (the rank condition would be violated if we estimated a fixed effects model). Estimates in column 1 of Table II are consistent qualitatively with the results reported in Table I and Figure 1: the wage gift increases worker effort levels early on in the experiment, but this effect dissipates over time.

### B. Fundraising Task

Table III provides a summary of the average total contributions collected per hour by treatment split into hourly time segments. Similar to the library task, in the beginning of the work day the *Gift* treatment yields significantly higher outcomes. For example, as shown in Figure 2, examining data in the first (prelunch) three hours of the capital campaign, we find an average total collection figure in the *Gift* treatment of \$11.00, whereas in the *noGift* treatment solicitors raised only \$6.40 per hour, a difference of 70 percent. A nonparametric Wilcoxon statistical test indicates that these averages are different at the  $p < 0.01$  level using a one-sided alternative. This outcome highlights the strength of the gift exchange effect.

Yet as Table III reveals, after a few hours on the job the average dollars collected (denoted "earnings") across the two treatments become quite similar. Indeed, only \$0.39 separates the average dollars collected per hour

TABLE II  
REGRESSION RESULTS<sup>a, c, d</sup>

Variable <sup>b</sup>	Library Task	Fundraiser	
		Hourly	3-Hour Block
<i>Gift</i>	10.9 (6.1)	3.4 (3.3)	13.8 (5.8)
<i>Time2</i>	-0.2 (2.0)	3.0 (2.8)	—
<i>Time3</i>	0.5 (2.0)	-3.6 (2.8)	—
<i>Time4</i>	-1.1 (2.0)	-1.2 (2.8)	—
<i>Time5</i>		2.6 (2.8)	—
<i>Time6</i>		-1.3 (2.8)	0.70 (4.4)
<i>Gift</i> × <i>Time2</i>	-6.6 (2.9)	2.0 (3.7)	—
<i>Gift</i> × <i>Time3</i>	-10.5 (2.9)	1.6 (3.7)	—
<i>Gift</i> × <i>Time4</i>	-10.2 (2.9)	-0.9 (3.7)	—
<i>Gift</i> × <i>Time5</i>		-5.8 (3.7)	—
<i>Gift</i> × <i>Time6</i>		-2.2 (3.7)	-12.6 (5.9)
<i>Constant</i>	40.7 (4.2)	6.6 (2.4)	19.2 (4.4)
<i>N</i>	76	138	46

<sup>a</sup>Standard errors are in parentheses beneath coefficient estimates.

<sup>b</sup>Dependent variable is the number of books logged in the library task and money raised in the fundraiser (in 3-hour blocks in column 3). *Gift* = 1 if the agent is in the gift treatment and 0 otherwise. *Time* variables are dichotomous, and equal 1 when in that period and 0 otherwise. *Gift* × *Time* variables are dichotomous interactions between the gift treatment and the time period.

<sup>c</sup>We also experimented with clustering standard errors and using a Tobit model; results are qualitatively identical. Column 1 empirical results are from *Stata*. Results summarized in columns 2 and 3 are from *Limdep*. The model in column 1 did not consistently converge in *Limdep* so we opted to present results from *Stata*. Nevertheless, the qualitative conclusions reached across statistical packages are identical in all cases (there are small deviations in the estimated standard errors).

<sup>d</sup>For the fundraising task, we also examined data on the number of solicitor contacts, dollars raised per contact, and the number of households approached. Focusing on the specification in column 3, for the number of solicitor contacts, we find that the coefficients and standard errors of *Gift*, *Time6*, *Gift* × *Time6*, and *Constant* are 3.8 (1.7), 1.8 (0.61), -1.41 (0.81), and 12 (1.3). For the dollars raised per contact specification, the respective coefficients and standard errors are 0.54 (0.45), -0.06 (0.36), -0.79 (0.48), and 1.66 (0.34). Finally, in the number of households approached specification, the coefficients and standard errors are 0.54 (4.4), 4.2 (1.0), 0.42 (1.4), and 34 (3).

TABLE III  
SUMMARY STATISTICS—AVERAGE EARNINGS BY TREATMENT<sup>a</sup>

	<i>NoGift</i>	<i>Gift</i>	Difference
Hour 1	6.6 (2.271)	10.00 (2.23)	-4.4
Hour 2	9.6 (3.572)	15.00 (3.95)	-5.4
Hour 3	3.00 (0.789)	8.00 (2.069)	-5.00**
Hour 4	5.40 (1.507)	7.846 (1.506)	-2.446
Hour 5	9.20 (2.08)	6.769 (0.856)	2.431
Hour 6	5.30 (2.547)	6.461 (1.483)	-1.161
Prelunch per hour (hours 1-3)	6.40 (1.803)	11.00 (1.443)	-4.6**
Postlunch per hour (hours 4-6)	6.633 (1.389)	7.026 (0.787)	-0.392
Entire day per hour (hours 1-6)	6.516 (1.474)	9.013 (0.814)	-2.496*
Number of solicitors	10	13	

<sup>a</sup> Cell observations give the mean and standard errors for the average earnings for solicitors in the nongift exchange and gift exchange treatments (columns 2 and 3, respectively). Column 4 provides the difference in average earnings between the solicitors in the nongift treatment and the earnings of solicitors in the gift treatments. The fourth column also indicates whether the difference is significant at the  $p < 0.05$  level (\*\*) or  $p < 0.10$  level (\*) using a nonparametric Wilcoxon test. For example, solicitors in the nongift treatment raised on average \$6.60 during the first hour of work, whereas solicitors receiving the gift earned an average of \$10.00 during this same hour.

after lunch across the two treatments. In addition, in any given hourly time period postlunch, the average total collection earnings across the two treatments are not statistically significant at conventional levels. The data summary in Table IV and Figure 2 provide an indication of why this convergence occurs. Although the solicitors in the nongift exchange treatment garnered hourly contributions in the three-hour block before lunch similar to those in the three-hour block after lunch (\$6.40 versus \$6.63), solicitors in the gift exchange treatment raised significantly less in the postlunch time period than they raised in the prelunch time period—\$7.03 per hour versus \$11 per hour. Using a matched pairs  $t$ -test, we find that this difference is statistically significant at the  $p < 0.05$  level using a two-sided alternative, whereas the temporal difference for the nongift exchange treatment is not significant.

Similar to the data entry task, this result suggests that this behavioral phenomenon is not restricted to a few select workers, rather it is more widespread. To make this point more clearly, we provide the individual level data aggregated by 3-hour blocks in Table V. Table V shows that in the *Gift* treatment nearly 70% of solicitors decrease their output whereas roughly 30% of solici-

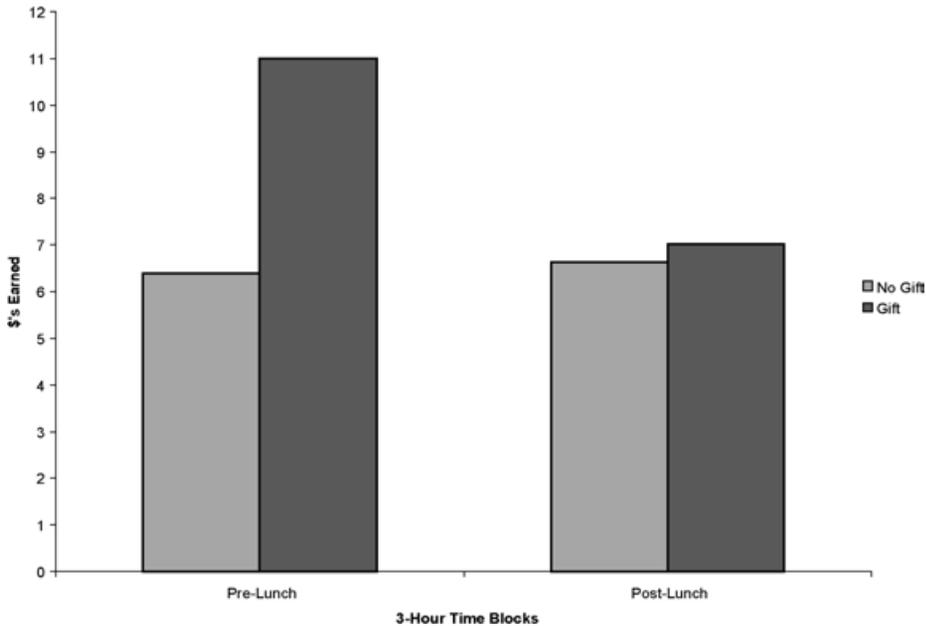


FIGURE 2.—Average earnings by 3-hour block.

tors increase their output. Alternatively, in the *noGift* treatment, these percentages are reversed: 70% of solicitors had increased output whereas only 30% had decreased output.<sup>8</sup> To offer an indication of the empirical significance of these differences, we provide Figure 3, which summarizes the individual mean

TABLE IV  
AVERAGE EARNINGS WITHIN TREATMENT BY 3-HOUR BLOCK<sup>a</sup>

	Prelunch	Postlunch	Difference
Gift exchange	11.00 (1.443)	7.026 (0.786)	3.974** (1.66)
Nongift exchange	6.40 (1.803)	6.633 (1.389)	-0.233 (1.29)

<sup>a</sup>Cell entries provide summary statistics for average earnings per hour for the 3-hour blocks before and after lunch across our two treatments (gift exchange and nongift exchange). Standard errors for the earnings are in parentheses. Column 4 provides the difference in average hourly earnings within a treatment across these blocks. \*\* indicates that the reported difference is statistically significant at the  $p < 0.05$  level using a matched pairs  $t$ -test.

<sup>8</sup>If we analyze the individual level data at the hourly block and compare the first and last hours, we find that in the *Gift* treatment nearly 70% of solicitors decrease their output whereas 30% of solicitors increase their output. Alternatively, in the *noGift* treatment half the solicitors had decreased output and half the solicitors had increased output.

TABLE V  
SUMMARY DATA—MONEY RAISED (IN 3-HOUR TIME BLOCKS)

	Participant Number	Pre Lunch	Post Lunch
<i>noGift</i>	1	6	7
	2	6	21
	3	20	24
	4	35	15
	5	6	25
	6	8	13
	7	0	4
	8	41	25
	9	49	51
	10	21	14
	Average	19.2	19.9
<i>Gift</i>	11	35	26
	12	32	34
	13	31	20
	14	14	19
	15	27	17
	16	42	25
	17	31	11
	18	26	3
	19	15	25
	20	42	16
	21	77	19
	22	29	33
	23	28	26
	Average	33	21.1

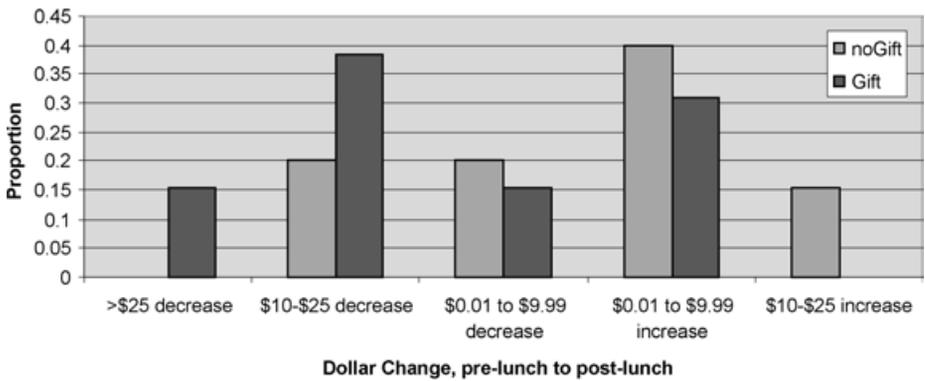


FIGURE 3.—Individual differences in mean contributions.

differences between the first and second three hours. The figure highlights the significant differences across the two treatments. In the *Gift* treatment more than half of our workers witnessed a decrease of more than \$10 whereas a majority of solicitors in the *noGift* treatment experienced increases in their overall earnings after lunch. For further evidence, we again turn to a panel data regression model in which we regress the individual monies raised on a dichotomous variable for treatment, dummy variables for time indicators, and their interaction. We again include random effects, and we should note that we experimented with the individual-specific variables found to be important in Landry, Lange, List, Price, and Rupp (2006) and found that their inclusion does not change the qualitative insights. Thus, for parsimony we exclude these variables. Columns 2 and 3 of Table II contain parameter estimates. As column 2 indicates, the hourly data are quite noisy and strong inference cannot be gained from these estimates. Yet the data split by three-hour blocks, pre- and postlunch, are consistent with the results discussed above (see column 3 of Table II): the wage gift worked well in the first few hours, but its influence waned considerably in the latter hours.

One interpretation of these findings is in the spirit of the psychology literature, which reminds us that there are differences between psychological processes in the short run and in the long run, or in the “hot” and “cold” phases (see, e.g., Loewenstein and Schkade (1999) and Loewenstein (2005)). For example, shortly after being injured, people with spinal injuries report very low quality of life. The same people report a much higher quality of life after a few years (Gilbert, Pinel, Wilson, Blumberg, and Wheatley (1998)). One way to think of this process is a change in the reference point. At the beginning of the process, people use their “old” reference point (being healthy, previous wage, etc.). With time, the reference point changes to account for the new environment (being injured, new wage, etc.) and thus behavior changes accordingly. In our case, an interpretation of our findings is that our agents’ effort levels may simply be adapting to new referentials in their progression from a hot to a cold phase of the time spent on their task, although in our case the hot phase lasts only a few hours. This would suggest that the equilibrium behavior is one where gift exchange has nominal long-term effects.

We must exercise caution when making this interpretation, however. The data show a pattern of convergence, but whether this result obtains because solicitors become physically exhausted as the day progresses is unknown. If our solicitors were to return the following morning and resumed the higher collection averages of the initial morning, we would wonder whether the physical nature of the task may play a role in the temporal effort profile.<sup>9</sup>

<sup>9</sup>As described in the notes to Table II, we also examined data on the number of solicitor contacts, dollars raised per contact, and the number of households approached. Data patterns (directionally) for the number of solicitor contacts and the dollars raised per contact are similar to those observed above: solicitors in the Gift treatment have a greater number of contacts and dollars

To examine the explanation for the higher solicitor performance in the *Gift* treatment in the early hours in greater detail, we dig a level deeper into the data. This is important in light of the fact that individual characteristics such as physical attractiveness and sociability might influence contributions (see Landry, Lange, List, Price, and Rupp (2006)). With this in mind, we collected additional data from a subset of solicitors who returned to work on the Sunday morning directly after their Saturday solicitation. Nine solicitors in the *Gift* treatment and four solicitors in the *noGift* treatment were included in the subset.

We provide Figure 4 to include these Sunday morning data. In this case, *Gift* subjects and *noGift* subjects perform similarly, because their outcomes are not statistically significant at conventional levels: *Gift* solicitors raised roughly \$6.50, whereas *noGift* solicitors raised nearly \$8.50. We view these data as evidence that exhaustion effects are not important in attenuating gift exchange in our Saturday afternoon data.

Most importantly for our purposes, as employers genuinely interested in creating a library at least cost and adequately funding a public good in North Carolina, we attempted to put a set of seminal findings from behavioral economics to work. Unfortunately, our plan backfired, because paying wages of merely \$10–12 would have netted more books logged and more donations collected per dollar spent on labor. Of course, similar to any empirical exercise, it is possible that a different experimental design and calibration, or other types of manipulations, might provide evidence that suggests certain behavioral find-

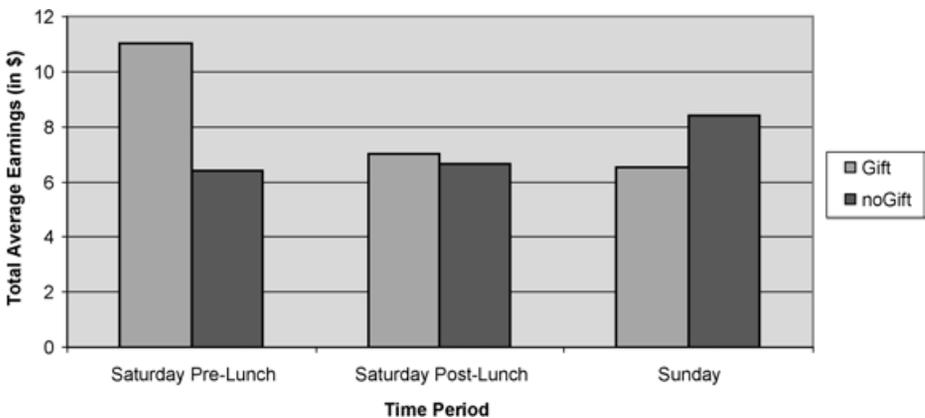


FIGURE 4.—Average solicitor earnings by 3-hour block.

raised per contact in the morning than solicitors in the *noGift* treatment, but the gaps lessen in the afternoon. Empirical results for the number of households approached show no discernable difference across treatment, before and after lunch.

ings have some relevancy for wage policies.<sup>10</sup> We hope that our study, which should be viewed as a first exploration of whether social preferences per se are enough to justify wage policies predicated on their existence, will stimulate further work using real pay and productivity measures from field settings of varying work durations.

#### 4. DISCUSSION

Besides their significance in testing important economic theories, our field results are important in interpreting empirical findings gathered in laboratory experiments. To our knowledge, there has been no direct test of whether experimental results gathered in the span of an hour or two can be used to make inference on tasks that are inherently much longer lived. There is, however, an emerging literature using laboratory and survey evidence that relates to our work.

Concerning laboratory evidence, a handful of recent studies explore conditions that facilitate or weaken the strength of gift exchange. For example, Fehr and List (2004) use a laboratory experiment to examine how chief executive officers in Costa Rica behave in sequential prisoner's dilemma games and compare their behavior with that of Costa Rican students. They find that for both subject pools, use of sanctions can reduce cooperative behavior (see also Gneezy and Rustichini (2000)). Houser, Xiao, McCabe, and Smith (2005) report similar results using a large sample of George Mason University undergraduate student subjects. Charness, Frechette, and Kagel (2004) report that laboratory gift exchange is considerably influenced by whether or not a comprehensive payoff table is made available to subjects. Likewise, data from Engelmann and Ortmann (2002) and Rigdon (2002) highlight that student behavior depends critically on parameterization and implementation considerations.

Although these studies represent important tests in drawing out the boundary conditions for laboratory gift exchange experiments, our results make a much different distinction. We interpret our findings as suggesting that great care should be taken before making inference from laboratory experiments, which might be deemed as hot decision making, to field environments, which typically revolve around cold decision making. Our results therefore lend insights to perhaps a different interpretation of these laboratory studies as well.<sup>11</sup>

<sup>10</sup>For instance, it would have been useful to obtain insights on how the subjects interpreted the wage increase in terms of a fairness variation. Yet, such a manipulation check would be unusual in this sort of environment and would potentially compromise the naturalness of the field experiment; thus we avoided such a design.

<sup>11</sup>Of course, with our data alone we cannot pinpoint whether the move to the field was necessary to observe the effects of task duration on behavior. We trust that future work will parse these factors and explore whether longer time frames in the lab can cause similar behavioral changes as

Survey evidence is another important source of data to study behavior in labor markets. In an extensive study of business executives, Bewley (1999) considers why wages are downwardly rigid during a recession. He reports that managers are worried that wage cuts might result in decreases in morale that would subsequently result in poor worker performance when the economy recovered, if not immediately. Put more succinctly, Bewley (1999, p. 54) argues that “many factors influence morale, including especially good personal contact with supervisors, a spirit of community within the business, and the perception that company policy is fair. Businesspeople value good morale because it fosters high productivity, low turnover, and ease in recruiting new workers.”

This line of reasoning highlights the importance of fairness considerations in cases of *negative reciprocity*. With respect to positive reciprocity, as considered in our study, Bewley’s evidence is less conclusive. He argues that morale is less important when considering wage increases, but finds that one main consideration when determining raises is the effect on employee turnover once the recession ends. In this spirit, our results are consonant with those of Bewley: his work suggests that there appears to be little connection between increasing pay and productivity, except to the extent that higher wages make it possible to attract, and retain, higher quality workers.

## 5. EPILOGUE

Empirical evidence shows that wages in labor markets do not always clear the market: in many cases, firms pay a higher than market-clearing wage, resulting in involuntary unemployment. One of the seminal theories that was put forward to explain this observation is the fair wage–effort theory, which predicts that wages above market-clearing levels can be an equilibrium in labor markets. Despite its profound implications, there does not exist compelling evidence from naturally occurring markets to support or refute this theory. This is not surprising in light of the difficulties associated with executing a clean empirical test of such phenomena. When such data are available, it is difficult to separate the consequences of factors of primary interest from the host of simultaneously occurring stimuli.

Experimental markets and laboratory studies alleviate many of these problems and provide an attractive basis for analyzing such issues. In this spirit, an influential line of laboratory experimental research has evolved that shows the

---

those observed herein. In this spirit, our study showcases the complementarities of field and lab experimentation that have been discussed in the literature (see, e.g., Harrison and List (2004)): given that we have discovered this behavioral pattern in naturally occurring environments, researchers might wish to return to the lab and detail the types of variables that can cause, attenuate, or exacerbate these duration effects. With this new evidence in hand, and perhaps equipped with deeper theoretical models, subsequent lab and field experiments can be conducted to examine the predictions.

importance of reciprocity in labor market settings, lending empirical support to the fair wage–effort theory.

Whether such results have implications for real labor markets remains an open empirical issue, however. We begin to resolve this uncertainty by exploring individual behavior in two distinct labor markets: data entry and door-to-door fundraising. We report two major insights. First, consistent with findings in the experimental literature, a higher wage was reciprocated by greater effort on the part of the employees during the early hours of the task. Second, this higher effort level was not persistent: after a few hours, effort levels in the gift treatment mirrored those in the nongift treatment.

More generally, a methodological contribution of this study is to show that field experiments can be used as a means to examine the *representativeness* of the environment. For example, before we can begin to make sound arguments that behavior observed in the lab is a good indicator of behavior in the field, we must explore whether certain dimensions of the laboratory environment might cause differences in behavior across these domains. This study highlights one of several important dimensions.<sup>12</sup> Future research should explore these insights more closely and extend the tests to examine other dimensions as well as negative reciprocity. We should stress that we do find behavioral similarities across the lab and the field over short durations (as does, e.g., Gneezy (2006)). A useful exercise for future research is to return to the lab to explore the robustness of our insights by examining, for example, whether and to what extent our results are robust to various lab manipulations.

*The Rady School of Management, University of California–San Diego, 9500 Gilman Dr., La Jolla, CA 92093, U.S.A.; [ugneezy@gsb.uchicago.edu](mailto:ugneezy@gsb.uchicago.edu)*  
and

*Dept. of Economics at Chicago, University of Chicago, 5807 South Woodlawn Ave., Chicago, IL 60637, U.S.A.; NBER, and RFF; [jlist@uchicago.edu](mailto:jlist@uchicago.edu).*

*Manuscript received August, 2005; final revision received May, 2006.*

#### APPENDIX: SOLICITATION SCRIPTS FOR DOOR-TO-DOOR FUNDRAISER

##### *ECU Center for Natural Hazards Mitigation Research—Script*

(If a minor answers the door, please ask to speak to a parent. Never enter a house.)

Hi, my name is \_\_\_\_\_. I am an **ECU student** visiting Pitt County households today on behalf of the newly formed *ECU Natural Hazards Mitigation Research Center*.

<sup>12</sup>Harrison and List (2004) provide a discussion of many other important dimensions.

*(Hand the blue brochure to the resident.)*

You may recall **hurricanes Dennis and Floyd** five years ago led to **widespread devastation** in Eastern North Carolina, hence the State authorized the new *Hazards Center*.

This research center will provide **support and coordination for research on natural hazard risks**, such as *hurricanes, tornadoes, and flooding*.

**The primary goal of the center is to reduce the loss of life and property damages due to severe weather events.**

**We are collecting contributions today on behalf of the new *ECU Hazards Center*. The Center is a nonprofit organization.**

**To raise funds for the new ECU Hazard Center we are conducting a charitable raffle:**

- **The winner receives a \$1,000 prepaid MasterCard.**
- **For every dollar you contribute, you will receive 1 raffle ticket.**
- **The odds of winning this charitable raffle are based on your contribution and total contributions received from other Pitt County households.**
- **The charitable raffle winner will be drawn at the Center on December 17th at noon. The winner will be notified and the results posted on the Center's web site.**
- **All proceeds raised by the raffle will fund the Hazards Center, which is a nonprofit organization.**

**Would you like to make a contribution today?**

*(If you receive a contribution, please write a receipt that includes their name and contribution amount.)*

If you have questions regarding the Center or want additional information, there is a phone number and web site address provided on the back of this blue brochure.

#### REFERENCES

- ADAMS, J. S. (1963): "Toward an Understanding of Inequality," *Journal of Abnormal and Social Psychology*, LXVII, 422–436. [1365]
- AGELL, J., AND P. LUNDBORG (1995): "Theories of Pay and Unemployment: Survey Evidence from Swedish Manufacturing Firms," *Scandinavian Journal of Economics*, 97, 295–307. [1366]
- AKERLOF, G. A. (1982): "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics*, 97, 543–569. [1365,1366]
- AKERLOF, G. A., AND J. L. YELLEN (1988): "Fairness and Unemployment," *American Economic Review, Papers and Proceedings*, 83, 44–49. [1365]

- (1990): “The Fair Wage-Effort Hypothesis and Unemployment,” *Quarterly Journal of Economics*, 105, 255–283. [1365]
- ARGYLE, M. (1981): *The Social Psychology of Work*. Baltimore, MD: Penguin Books. [1366]
- BERG, J., J. W. DICKHAUT, AND K. A. MCCABE (1995): “Trust, Reciprocity, and Social History,” *Games and Economic Behavior*, 10, 122–142. [1366]
- BEWLEY, T. F. (1999): *Why Wages Don't Fall During a Recession*. Cambridge, MA: Harvard University Press. [1380]
- CHARNESS, G. (2004): “Attribution and Reciprocity in an Experimental Labor Market,” *Journal of Labor Economics*, 22, 665–688. [1366]
- CHARNESS, G., G. R. FRECHETTE, AND J. H. KAGEL (2004): “How Robust Is Laboratory Gift Exchange?” *Experimental Economics*, 7, 189–205. [1379]
- ENGELMANN, D., AND A. ORTMANN (2002): “The Robustness of Laboratory Gift Exchange: A Reconsideration,” Working Paper, Charles University. [1379]
- FALK, A., AND S. GÄCHTER (2002): “Reputation and Reciprocity—Consequences for Labour Relations,” *Scandinavian Journal of Economics*, 104, 1–26. [1366]
- FEHR, E., AND A. FALK (1999): “Wage Rigidity in a Competitive Incomplete Contract Market,” *Journal of Political Economy*, 107, 106–134. [1366]
- FEHR, E., AND S. GÄCHTER (2000): “Fairness and Retaliation: The Economics of Reciprocity,” *Journal of Economic Perspectives*, 14, 159–181. [1366]
- FEHR, E., S. GÄCHTER, AND G. KIRCHSTEIGER (1997): “Reciprocity as a Contract Enforcement Device,” *Econometrica*, 65, 833–860. [1366]
- FEHR, E., AND J. A. LIST (2004): “The Hidden Costs and Returns of Incentives—Trust and Trustworthiness among CEOs,” *Journal of the European Economic Association*, 2, 743–771. [1379]
- FEHR, E., G. KIRCHSTEIGER, AND A. RIEDL (1993): “Does Fairness Prevent Market Clearing? An Experimental Investigation,” *Quarterly Journal of Economics*, 108, 437–460. [1366]
- GILBERT, D. T., E. C. PINEL, T. D. WILSON, S. J. BLUMBERG, AND T. WHEATLEY (1998): “Immune Neglect: A Source of Durability Bias in Affective Forecasting,” *Journal of Personality and Social Psychology*, 75, 617–638. [1367,1377]
- GNEEZY, U. (2006): “Bonuses versus Predetermined Wages: A Study of Incentives and Reciprocity,” Mimeo, University of Chicago. [1381]
- GNEEZY, U., AND A. RUSTICHINI (2000): “A Fine Is a Price,” *Journal of Legal Studies*, XXIX, 1–18. [1379]
- GUTH, W. (1995): “On Ultimatum Bargaining Experiments: A Personal Review,” *Journal of Economic Behavior and Organization*, 27, 329–344. [1366]
- GUTH, W., R. SCHMITTBERGER, AND B. SCHWARZE (1982): “An Experimental Analysis of Ultimatum Bargaining,” *Journal of Economic Behavior and Organization*, 3, 367–388. [1366]
- HARRISON, G. W., AND J. A. LIST (2004): “Field Experiments,” *Journal of Economic Literature*, XLII, 1009–1055. [1380,1381]
- HOUSER, D., E. XIAO, K. MCCABE, AND V. SMITH (2005): “When Punishment Fails: Research on Sanctions, Intentions and Non-Cooperation,” Working Paper, George Mason University. [1379]
- KAHNEMAN, D., J. L. KNETSCH, AND R. H. THALER (1986): “Fairness as a Constraint on Profit Seeking: Entitlements in the Market,” *American Economic Review*, 76, 728–741. [1366]
- KATZ, L. F. (1986): “Efficiency Wage Theories: A Partial Evaluation,” Working Paper 1906, NBER. [1368]
- LANDRY, C., A. LANGE, J. A. LIST, M. K. PRICE, AND N. RUPP (2006): “Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment,” *Quarterly Journal of Economics*, 121, 747–782. [1369,1377,1378]
- LOEWENSTEIN, G. (2005): “Hot–Cold Empathy Gaps and Medical Decision–Making,” *Health Psychology*, 24, S49–S56. [1366,1377]
- LOEWENSTEIN, G., AND D. SCHKADE (1999): “Wouldn't It Be Nice? Predicting Future Feelings,” in *Well Being: The Foundations of Hedonic Psychology*, ed. by D. Kahneman, E. Diener, and N. Schwartz. Russell Sage Foundation. [1377]

- RIGDON, M. (2002): "Efficiency Wages in an Experimental Labor Market," *Proceedings of the National Academy of Sciences*, 99, 13348–13351. [1379]
- ROTH, A. (1995): "Bargaining Experiments," in *The Handbook of Experimental Economics*, ed. by J. Kagel and A. Roth. Princeton, NJ: Princeton University Press, Chap. 4, 253–348. [1366]
- SOBEL, J. (2005): "Interdependent Preferences and Reciprocity," *Journal of Economic Literature*, 43, 392–436. [1366]
- STEERS, R. M., AND L. W. PORTER (1991): *Motivation and Work Behavior* (Fifth Ed.). New York: McGraw-Hill. [1366]